VerA: Versatile Anonymization Applicable to Clinical Facial Photographs

Majed El Helou*[1], Doruk Cetin*[2], Petar Stamenkovic[1], Niko Benjamin Huber[2], Fabio Zünd[1]
[1]Media Technology Center, ETH Zürich, Switzerland, [2]Align Technology Zürich, Switzerland

WACV 2025 TUCSON, ARIZONA • FEB 28 - MAR 4

Motivation

- **Clinical anonymization:** Semantic preservation of selected areas to illustrate medical intervention
- **Paired anonymization:** Anonymized images with consistent identity across multiple photographs

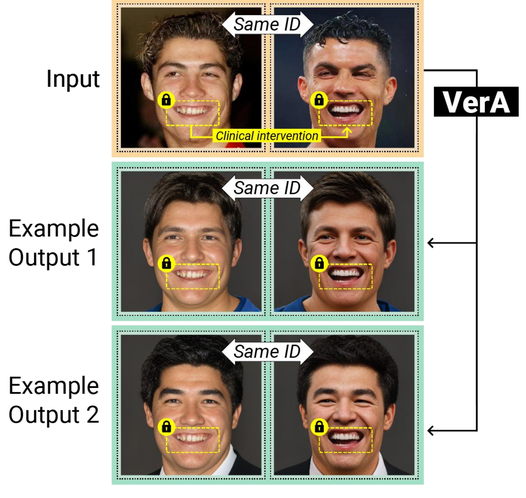
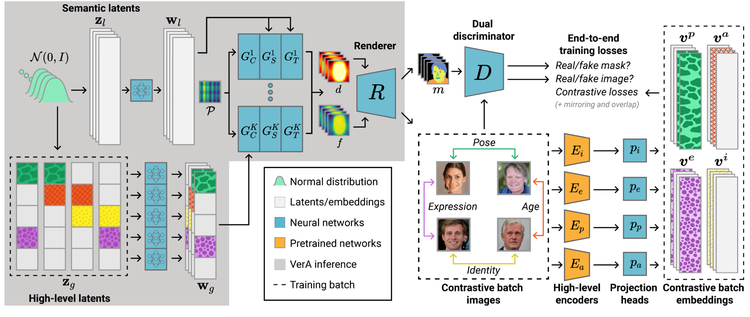


Fig. High-level comparison with common methods

Method Year & Ref.	CIAGAN '20 [42]	FIT '20 [18]	DP2 '23 [24]	RiDDLE '23 [36]	FALCO '23 [7]	IDeudemon '23 [66]	Ours -
Generation	CIAGAN	FIT	DP2	StyleGAN2		GFP-GAN	Ours
De-ID	✓	✓	✓	✓	✓	✓	✓
Photorealism	✗	✗	✓	✗	✗	✓	✓
In-place anon	✗	✗	✓	✗	✗	✓	✓
Clinical anon	✗	✗	✗	✗	✗	?	✓
Paired anon	✗	✗	✗	✗	✗	✗	✓
Clinical pair	✗	✗	✗	✗	✗	✗	✓

Methodology

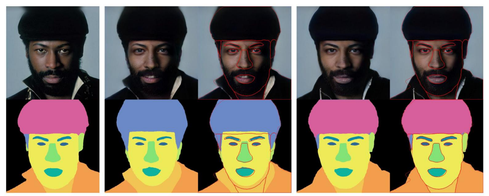
Fig. Model architecture and training pipeline



- **Explicit high-level and semantic control:** generator with dual latent space, trained via contrastive learning

Fig. Accumulated semantic and high-level edits

- **Specialized inversion:** joint optimization of multiple input and additional segmentation loss during inversion

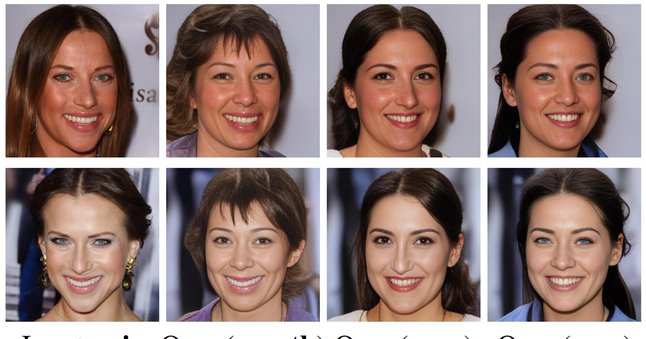


(a) Input (b) Inverted w/o guidance (c) Inverted w/ guidance

Fig. The effect of segmentation loss on inversion

Results

Fig. Clinical, paired anonymization results

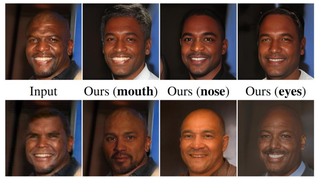


Input **pair** Ours (**mouth**) Ours (**nose**) Ours (**eyes**)

Fig. Standard, paired anonymization results

Input CIAGAN [42] FIT [18] DP2 [24] RiDDLE [36] FALCO [7] Ours

Fig. Clinical, single-image anonymization results



Input Ours (**mouth**) Ours (**nose**) Ours (**eyes**)

FIT [18] DP2 [24] RiDDLE [36] FALCO [7]

Fig. Standard anonymization results on full-scale images

Tab. De-identification rates

	Face Recog. ↑ Setting	CASIA [73] Single	CASIA [73] Paired	VGGFace2 [9] Single	VGGFace2 [9] Paired	FaceNet [54] Single	FaceNet [54] Paired
Standard	CIAGAN [42]	98.6	97.8	98.4	97.4	**99.9**	**100.0**
	FIT [18]	99.1	98.2	98.3	98.8	99.1	99.8
	DP2 [24]	99.1	99.2	98.5	98.8	83.2	85.0
	RiDDLE [36]	**99.9**	**100.0**	**99.7**	**99.6**	96.7	96.8
	FALCO [7]	98.7	98.6	98.7	**99.6**	90.6	90.4
	Ours	98.6	98.4	95.8	96.0	97.9	98.2
Clinical	Ours (mouth)	99.8	99.8	98.9	99.0	98.2	98.0
	Ours (nose)	99.8	99.4	96.1	95.6	98.9	99.6
	Ours (eyes)	99.8	99.8	98.8	99.2	99.6	**100.0**

Tab. Re-identification rates within pairs

	Face Recog. Setting	CASIA [73] Rate ↑	CASIA [73] Dist. ↓	VGGFace2 [9] Rate ↑	VGGFace2 [9] Dist. ↓	FaceNet [54] Rate ↑	FaceNet [54] Dist. ↓
	Input pair	92.8	-	90.0	-	89.20	-
Single image	CIAGAN [42]	65.4	+0.161	74.1	+0.136	74.5	+1.694
	FIT [18]	15.6	+0.355	10.8	+0.443	24.4	+4.771
	DP2 [24]	30.8	+0.293	23.6	+0.347	75.6	+0.976
	RiDDLE [36]	12.0	+0.469	10.4	+0.515	29.6	+4.939
	FALCO [7]	52.8	+0.224	40.4	+0.300	62.0	+2.413
	Ours (standard)	47.2	+0.242	26.8	+0.364	69.6	+1.702
Paired	Ours (standard)	85.2	+0.065	71.6	+0.154	85.6	+0.132
	Ours (mouth)	89.6	+0.026	78.8	+0.102	87.2	+0.109
	Ours (nose)	**90.0**	**-0.001**	**82.4**	**+0.074**	**89.2**	**+0.065**
	Ours (eyes)	88.8	+0.030	80.4	+0.102	83.6	+0.324

Tab. Semantic preservation

	Method	ℓ_1 distance ↓ Mouth	 Nose	 Eyes	PSNR ↑ Mouth	 Nose	 Eyes	Semantic IoU ↑ Mouth	 Nose	 Eyes	Landmark offset ↓ Mouth	 Nose	 Eyes
Standard	CIAGAN [42]	34.25	25.97	52.74	15.08	17.36	11.65	0.55	0.58	0.02	16.84	17.17	40.97
	FIT [18]	20.88	15.38	29.48	19.26	21.65	16.23	0.78	0.84	0.64	8.40	8.07	7.97
	DP2 [24]	41.03	30.96	53.73	13.47	15.55	11.38	0.51	0.60	0.27	31.37	33.57	27.75
	RiDDLE [36]	33.26	28.54	41.56	15.30	16.55	13.57	0.71	0.76	0.59	13.32	19.08	9.97
	FALCO [7]	32.71	26.61	40.57	15.52	17.47	13.60	0.67	0.79	0.55	15.51	15.27	10.60
	Ours	34.20	18.82	41.42	14.70	19.25	13.20	0.66	0.80	0.60	16.25	14.24	9.49
Clinical	Ours (mouth)	**0.22**	18.86	43.21	**51.06**	19.30	12.94	**0.92**	0.81	0.57	**6.81**	14.17	10.18
	Ours (nose)	33.94	**0.17**	42.51	14.76	**51.33**	13.04	0.67	0.94	0.59	16.38	**5.67**	9.63
	Ours (eyes)	34.58	18.90	**0.48**	14.64	19.28	**40.82**	0.65	0.80	**0.83**	17.61	14.70	**6.53**

Tab. Downstream utility

Method	FID ↓ FFHQ	FID ↓ CelebAHQ	Bounding box ↑ MTCNN	Bounding box ↑ Dlib	Face detection ↑ MTCNN	Face detection ↑ Dlib
CIAGAN [42]	138.87	74.91	0.82	0.91	0.93	0.94
FIT [18]	125.67	79.80	**0.92**	**0.97**	0.97	0.99
DP2 [24]	69.36	19.34	0.88	0.88	0.92	0.96
RiDDLE [36]	137.96	68.63	0.90	0.95	**1.00**	**1.00**
FALCO [7]	87.81	37.62	0.90	0.94	0.99	**1.00**
Ours	**57.77**	**12.57**	0.91	0.95	0.96	0.99

- [7] Barattin et al. "Attribute-preserving Face Dataset Anonymization via Latent Code Optimization" CVPR 2023
- [9] Cao et al. "VGGFace2: A dataset for recognising faces across pose and age" FG 2018
- [18] Gu et al. "Password-conditioned anonymization and deanonymization with face identity transformers" ECCV 2020
- [24] Hukkelås et al. "DeepPrivacy2: Towards Realistic Full-Body Anonymization" WACV 2023
- [36] Li et al. "RiDDLE: Reversible and Diversified De-identification with Latent Encryptor" CVPR 2023
- [42] Maximov et al. "CIAGAN: Conditional identity anonymization generative adversarial networks" CVPR 2020
- [54] Schroff et al. "FaceNet: A unified embedding for face recognition and clustering" CVPR 2015
- [66] Wen et al. "Divide and Conquer: a Two-Step Method for High Quality Face De-identification with Model Explainability" ICCV 2023
- [73] Yi et al. "Learning face representation from scratch" arXiv 2014